

DESPUÉS DE COST/ES0601: COMPARACIÓN AUTOMÁTICA DE MÉTODOS DE HOMOGENEIZACIÓN DE SERIES

José A. GUIJARRO
Agencia Estatal de Meteorología, España
jguijarrop@aemet.es

RESUMEN

La Acción COST ES0601 “*Advances in Homogenisation Methods of Climate Series: An Integrated Approach (HOME)*” ha permitido realizar un exhaustivo trabajo de intercomparación de métodos de homogeneización que ha involucrado a investigadores de una veintena de países. Sin embargo, la complejidad del problema planteado ha hecho que sólo los métodos automáticos pudieran entregar la totalidad de las soluciones, haciendo difícil la comparación de los resultados. Por otra parte, esta Acción ha estimulado la mejora de muchos paquetes de homogeneización, pero repetir la intercomparación de la misma manera resulta poco práctico por la carga de trabajo que implica. Como alternativa se puede plantear una comparación automática, y aquí se presentan y discuten los resultados de aplicar varios paquetes de homogeneización a 1500 series simuladas de temperatura mensual con uno o dos saltos en la media. Exceptuando los malos resultados de las homogeneizaciones absolutas, los errores de los demás métodos no son demasiado dispares, y resulta interesante constatar que las tendencias de las series reconstruidas no presentan ningún sesgo apreciable, lo que aumenta la confianza en su uso para la evaluación de la variabilidad climática.

Palabras clave: Métodos Homogeneización Series, Comparación Automática.

ABSTRACT

The COST Action ES0601 “*Advances in Homogenisation Methods of Climate Series: An Integrated Approach (HOME)*” allowed researchers of about 20 countries to perform an exhaustive study of inter-comparison of homogenization methods. Yet, due to the complexity of the benchmark database, only automatic methods could provide a complete set of solutions, making difficult the comparison of the results. On the other hand, this Action has stimulated the improvement of many homogenization packages, but repeating the inter-comparison in the same way is not foreseen because of the high amount of work involved. The alternative approach is to perform automatic comparisons, and here the results of applying several homogenization programs to 1500 synthetic monthly temperature series with one or two shifts in the mean are presented and discussed. Taking apart the bad results of the absolute homogenizations, the other are not too different. Moreover, it is interesting to note that the trends of the corrected series do not present significant bias, hence increasing their reliability for climate variability studies.

Key words: Series Homogenization Methods, Automatic Benchmarking.

1. INTRODUCCIÓN

Las series climatológicas no sólo reflejan las variaciones del clima de la zona en la que están situados los observatorios, sino que también se ven afectadas por alteraciones locales (reubicaciones, cambios en el entorno) y por errores de calibración o cambios en los métodos de observación. Por ello se han venido desarrollando métodos para tratar de homogeneizar las series, esto es, eliminar las influencias perturbadoras que enmascaran en mayor o menor grado la verdadera evolución del clima. Peterson *et al.* (1998) y Aguilar *et al.* (2003) presentan sendas revisiones de los mismos, que habitualmente tenían que ser programados por cada investigador. Algunos de ellos, sin embargo, han ofrecido esos programas para su uso por la comunidad climatológica, aunque se hacía necesario algún tipo de estudio que evaluara sus resultados.

Este estudio es el que se ha abordado en la Acción COST ES0601 “*Advances in Homogenisation Methods of Climate Series: An Integrated Approach (HOME)*”, en la que durante cinco años (2007-2011) se ha llevado a cabo un exhaustivo trabajo de intercomparación de métodos de homogeneización que ha involucrado a investigadores de una veintena de países, y que ha consistido en la generación de un conjunto de series problema que los participantes debían homogeneizar sin conocer la solución. Posteriormente se ha desvelado ésta y se han estudiado los errores de cada contribución, analizando los puntos fuertes y débiles de cada metodología. Un extenso resumen de todos estos trabajos se puede consultar en Venema *et al.* (2011).

Sin embargo, durante el desarrollo de estos trabajos se han puesto de manifiesto algunos inconvenientes. Por un lado, el conjunto de series problema era muy complejo, pues estaba compuesto de 80 redes simuladas de observación (40 de temperatura y 40 de precipitación) con datos mensuales de entre 5 y 15 series cada una con 100 años (1200 datos) de longitud. Esto ha producido una gran disparidad en el volumen de soluciones aportadas por cada metodología, pues las de funcionamiento automático pudieron homogeneizar la totalidad de las series, mientras que los métodos manuales más costosos sólo ofrecieron resultados para unas pocas. Y, por otro lado, las inhomogeneidades artificialmente introducidas en las series problema consistían en múltiples saltos en la media de posición aleatoria y magnitud normalmente distribuida, así como en tendencias locales de diverso tamaño y longitud, lo que puso de manifiesto la disparidad de criterio entre las homogeneizaciones drásticas (homogeneizar cualquier salto sospechoso) y las conservadoras (elevar el umbral de detección para evitar falsos positivos). Si a esto añadimos que algunos métodos automáticos presentaron errores debidos no a la metodología empleada sino al procesamiento de las soluciones para presentarlas en la complicada estructura exigida, se comprenderá que la comparación de los distintos resultados ha sido un problema en sí misma, al no poderse basar en un conjunto homogéneo de soluciones.

Por otra parte, muchas de las metodologías utilizadas han continuado su evolución y mejora, en gran parte gracias a las fructíferas discusiones mantenidas en el marco de esta Acción COST. Por tanto sería conveniente repetir el ejercicio de intercomparación con las nuevas versiones, pero el trabajo implicado lo hace poco viable en un futuro inmediato. La alternativa consiste en efectuar las comparaciones de forma automática, lo que permite realizar las pruebas un gran número de veces y minimizar así los efectos aleatorios. El inconveniente es que esto sólo es aplicable a los métodos que puedan funcionar sin intervención humana alguna, pero son precisamente estos métodos los de mayor interés en climatología operativa, puesto que las bases de datos se van ampliando año tras año, haciendo necesario repetir la homogeneización de sus series con cierta frecuencia. Los métodos

manuales, más laboriosos, podrán seguir usándose si se desea en estudios que involucren a un reducido número de series (normalmente seleccionadas por su mayor longitud).

Aquí se presenta un primer ejercicio de comparación automática de paquetes de homogeneización, detallando cómo se han construido las series problema y cómo se ha realizado la homogeneización, analizando los resultados obtenidos y planteando futuras comparaciones.

2. METODOLOGÍA

La complejidad del banco de datos de prueba usado en HOME, con multitud de saltos en la media y tendencias locales de magnitud y posición aleatorias, dificultó el análisis de los distintos aspectos de la homogeneización, como la sensibilidad de la detección, la frecuencia de falsas detecciones, los errores en la localización, y la correcta evaluación de las correcciones a aplicar.

Por ello en este primer experimento de comparación automática se ha planteado un caso sencillo, simulando redes termométricas en las que hay cinco series homogéneas que pueden servir de referencia y otras cinco que tienen uno o dos saltos en la media de $\pm 2^{\circ}\text{C}$, magnitud que ningún método debería tener dificultad en detectar. De este modo la verificación se puede centrar en evaluar los errores que introduce la corrección de las series inhomogéneas en el cálculo de sus promedios, tendencias y variabilidad.

Se ha comenzado por generar 100 ubicaciones aleatorias sobre un área de 4° de longitud por 3° de latitud y altitudes entre 600 y 900 m. Para la primera de ellas se han simulado 12 series mensuales de 60 términos (años) cada una, con una desviación típica de $1,5^{\circ}\text{C}$ y las medias mensuales de la tabla 1, correspondientes a las temperaturas máximas diarias medias de 53 estaciones de la zona central de la cuenca del Duero.

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
TA	7,9	10,5	14,2	16,3	20,6	25,8	29,9	29,3	25,2	18,7	12,2	8,3

TABLA 1: Medias mensuales de temperatura máxima (TA, en $^{\circ}\text{C}$) usadas para simular las series de datos de prueba.

A la localidad más próxima a esta primera se le asignan los mismos valores incrementados en valores aleatorios de distribución normal con media nula y desviación típica igual a 1,5 multiplicados por un factor de 0,20. Los datos de las otras 98 estaciones se generaron de la misma manera, a partir de los datos de la estación más próxima a cada una de ellas. Esto constituyó una primera red básica de estaciones, denominada TA20, a la que se añadieron otras dos (TA40 y TA80) construidas del mismo modo pero usando factores de 0,40 y 0,80 respectivamente al añadir las perturbaciones aleatorias. Las últimas intervenciones sobre estas series fueron: a) Ajustar las temperaturas según la altitud asignada a cada estación; b) Variar la oscilación termométrica anual en un valor aleatorio entre $\pm 20\%$; y c) Añadir una tendencia constante de 2°C/siglo a todas las series. Estas tres redes básicas se caracterizan por tener distintos grados de correlación entre las estaciones, elevadas, moderadas y bajas, como se observa en la figura 1.

A partir de cada uno de estos tres conjuntos de datos básicos se simularon 100 conjuntos de series, procediendo del siguiente modo:

- 1. Se seleccionan 40 estaciones aleatoriamente de las 100 disponibles (figura 2).
- 2. Se alteran las primeras 5 series introduciendo saltos de $\pm 2^{\circ}\text{C}$: dos en posiciones fijas en las tres primeras, y sólo una, pero con posición aleatoria, en las series cuarta y quinta, manteniendo siempre inalterados los últimos 10 años (fig.3).
- 3. Las estaciones 6 a 10 no se modifican, para que puedan servir como referencias homogéneas.
- 4. A las siguientes 30 estaciones se les anula los datos de entre 30 y 50 años (del 50 al 83,3% de ellos), con inicio y final aleatorios, para simular las estaciones de corto periodo de observación que habitualmente existen en las redes reales.

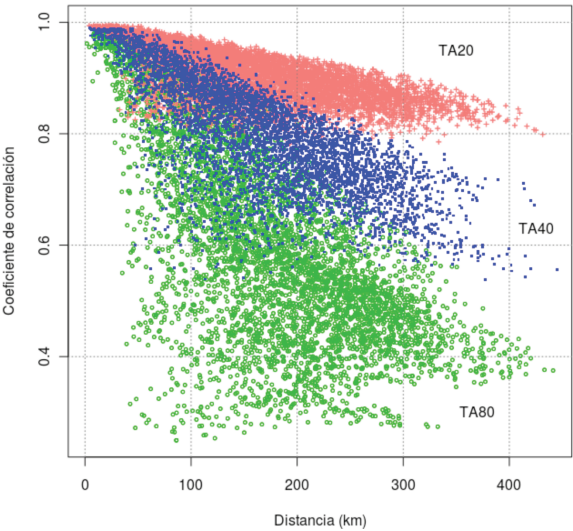


FIG. 1: Correlograma de las primeras diferencias de las series de las tres redes básicas, TA20, TA40 y TA80, mostrando cómo disminuyen las correlaciones entre sus estaciones.

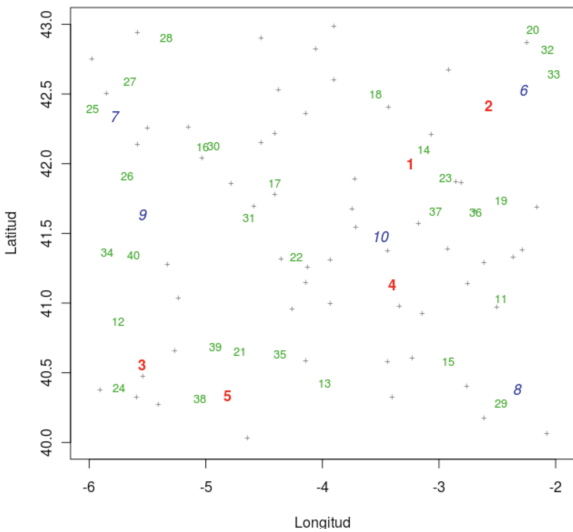


FIG. 2: Ejemplo de red problema. Las 40 estaciones numeradas se seleccionan aleatoriamente de entre las 100 localidades posibles. Las primeras 5 se inhomogeneizan, y las 30 últimas se blanquean dejándoles únicamente de 10 a 30 años (del 16,6 al 50%) de datos.

Los programas y modalidades de homogeneización que se han probado han sido los siguientes:

1. ACMANT: **Acm** (Domonkos, 2011).
2. Climatol: Aplicado a 10, 20 y 40 estaciones con correcciones constantes, **cl1**, **cl2**, **cl4**, y variables, **C11**, **C12**, **C14** (Guijarro, 2011).
3. RHTestV3: **RHa**, **RHA**, **RHr**, **RHR**, donde la última letra hace referencia a homogeneizaciones absolutas (a, A) o relativas (r, R), y a si se ajusta sólo la media (a, r) o también las frecuencias por cuantiles (A, R). (Wang y Feng, 2010).
4. HOMER: **Hom**, con la siguiente secuencia de operaciones: *Pairwise detection*, *Joint detection*, y *Correction*. (Programa generado durante la Acción COST ES0601, todavía sin publicar).
5. USHCN_v52d: Aplicado a 10, 20 y 40 estaciones, **US1**, **US2** y **US4** (Menne y Williams, 2009).

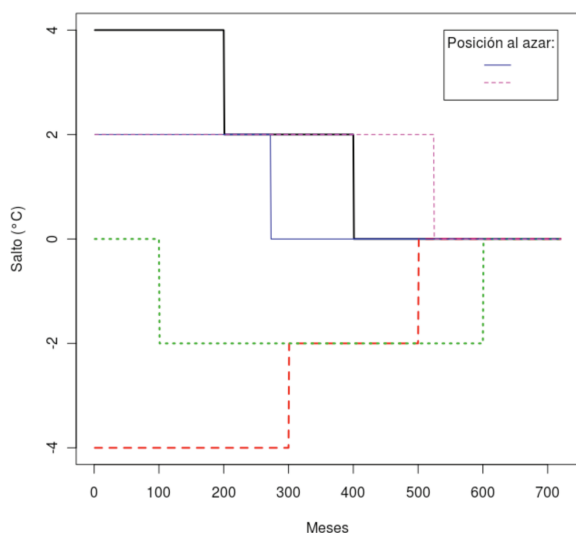


FIG. 3: *Inhomogeneidades aplicadas a las primeras 5 series. (2 saltos fijos a las tres primeras, y uno sólo, pero con posición aleatoria, a las otras dos).*

En general estos programas se aplicaron sobre las 10 series completas de cada red generada (5 muy inhomogéneas y 5 homogéneas), solicitando como resultado las series homogeneizadas y ajustadas para el último periodo homogéneo. La excepción fue RHTestV3, que no construye las series de referencia automáticamente, por lo que para cada serie inhomogénea se le proporcionó la referencia homogénea más cercana (lo que hay que tener en cuenta a la hora de comparar sus resultados con los de los otros programas, que deben buscar la mejor referencia por sí mismos). Además, los programas que toleraban series con una gran proporción de datos ausentes (Climatol y USHCN_v52d) se probaron también proporcionando 10 y 30 series incompletas adicionales, para evaluar el valor añadido por las mismas para la mejora de las homogeneizaciones.

Todo el proceso se implementó en un programa escrito en R (R Development Core Team, 2010) sobre un PC con S.O. Linux (Lubuntu 11.04), por lo que ACMANT, que está escrito para Windows, se ejecutó mediante la utilidad *wine* de Linux. Otros dos paquetes escritos para Windows, MASH y AnClim, se intentaron aplicar de la misma forma, pero sin éxito, por lo que su comparación con el resto de programas es una tarea pendiente para el futuro. Otro programa problemático fue USHCN_v52d, suministrado en código fuente Fortran, que dio errores de compilación. Para

solventarlos tuvo que ser trasladado a un PC con la versión de Linux utilizada por los desarrolladores (RedHat EL4).

RESULTADOS Y DISCUSIÓN

En cada una de las pruebas se han comparado las cinco series homogeneizadas por los programas con sus valores originales antes de introducirles las inhomogeneidades problema, calculando los siguientes parámetros:

1. Los errores típicos (RMSE), como medida general de la bondad de la homogeneización.
2. Los errores de las tendencias, importantes para la detección del cambio climático.
3. Los errores de las medias, para aplicaciones tales como la cartografía climática.
4. Los errores de las desviaciones típicas, potencialmente útiles en estudios de variabilidad (aunque para estas aplicaciones es preferible usar datos diarios).

Los resultados de los dos primeros pueden verse en los diagramas de caja de la figura 4, desglosados para las redes con buenas, regulares y malas correlaciones. Cada caja de los gráficos presentados resume 500 valores (5 series problema en 100 pruebas diferentes), y se ha elegido una escala constante para facilitar la comparación visual, si bien en los peores casos hay resultados que se salen de los gráficos.

Podemos ver que, en general, todos los programas utilizados mejoran los resultados de las series inhomogéneas problema (Inh), como cabía esperar dado el elevado grado de inhomogeneidad impuesto a las mismas. Así, los errores típicos son mayoritariamente menores que antes de la homogeneización, y las tendencias no presentan sesgos apreciables ni en los casos con mayor dispersión de resultados. También se pone de manifiesto que las menores correlaciones de las redes TA40 y TA80 incrementan la dificultad del problema, y esto sucede incluso en el caso de las homogeneizaciones absolutas RHa y RHA, a pesar de que no utilizan ninguna serie de referencia, porque las series TA40 y TA80 también presentan una mayor variabilidad, reduciendo así la relación señal/ruido y por tanto la capacidad de detección de los saltos. Por consiguiente, la homogeneización absoluta debe evitarse por todos los medios, siendo aconsejable procurarse series de referencia aunque sea a partir de variables *proxy* si no son posibles otras alternativas.

Los demás resultados proceden de homogeneizaciones relativas y, aparte de RHR, dan resultados bastante similares. Los peores resultados de RHR proceden de la corrección por cuantiles, que también dio malos resultados en el banco de pruebas de la Acción COST ES0601 (Venema *et al.*, 2012). Los otros resultados de RHTestV3 (RHR), que sólo corrigen la media, son mucho mejores (como cabe esperar si consideramos que las series problema tienen saltos constantes), igualando o superando a los demás, si bien hay que recordar que a este método se le ha proporcionado la referencia homogénea más próxima a cada serie problema, mientras que los demás programas han tenido que construirse las referencias de cada serie con sus propios algoritmos a partir de las demás (incluyendo las inhomogéneas).

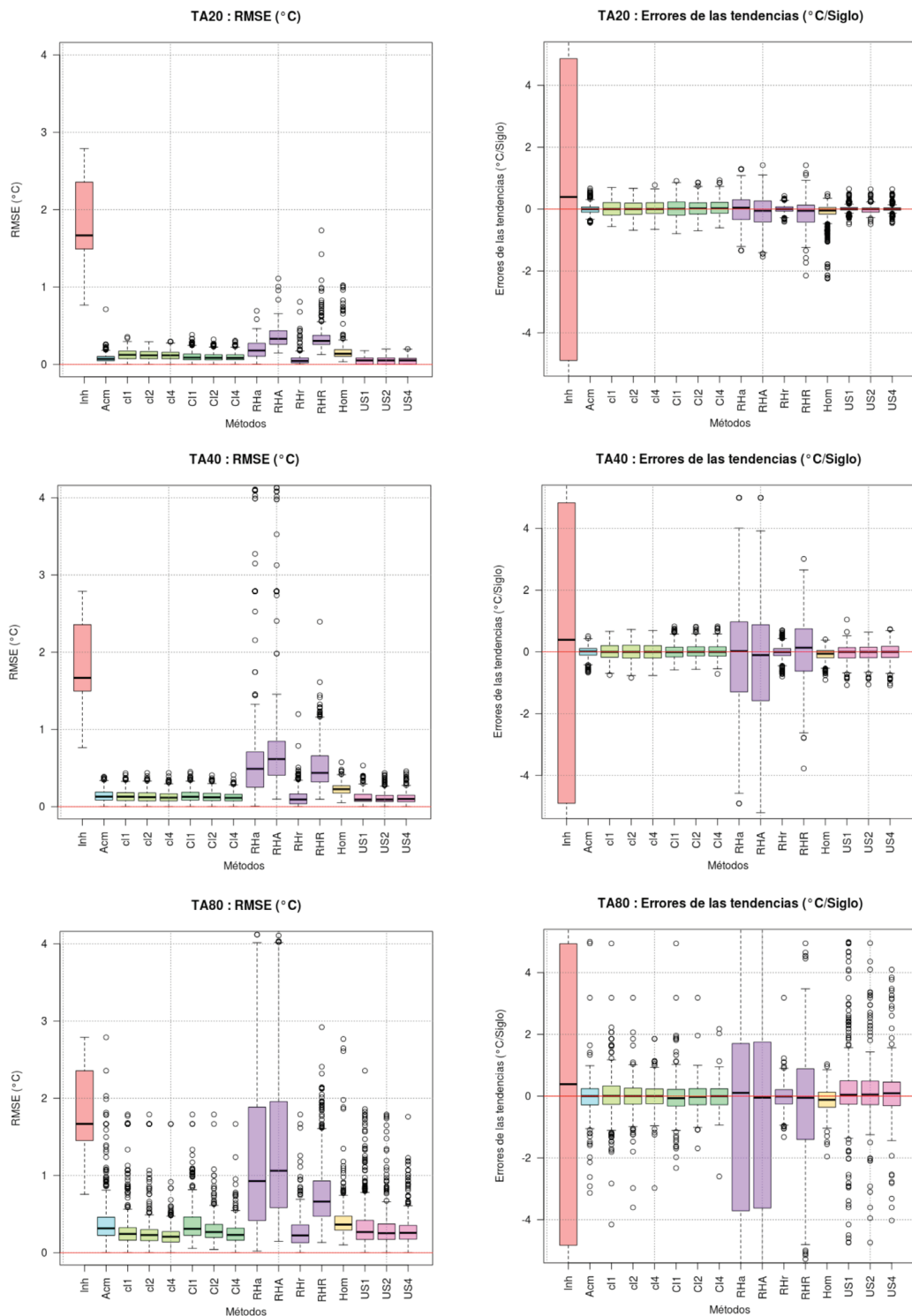


FIG. 4: Errores típicos RMSE (izquierda) y errores de las tendencias (derecha) de las series problema, antes (Inh) y después de ser homogeneizadas por los programas ensayados. (De arriba a abajo, redes con correlaciones buenas, regulares y malas).

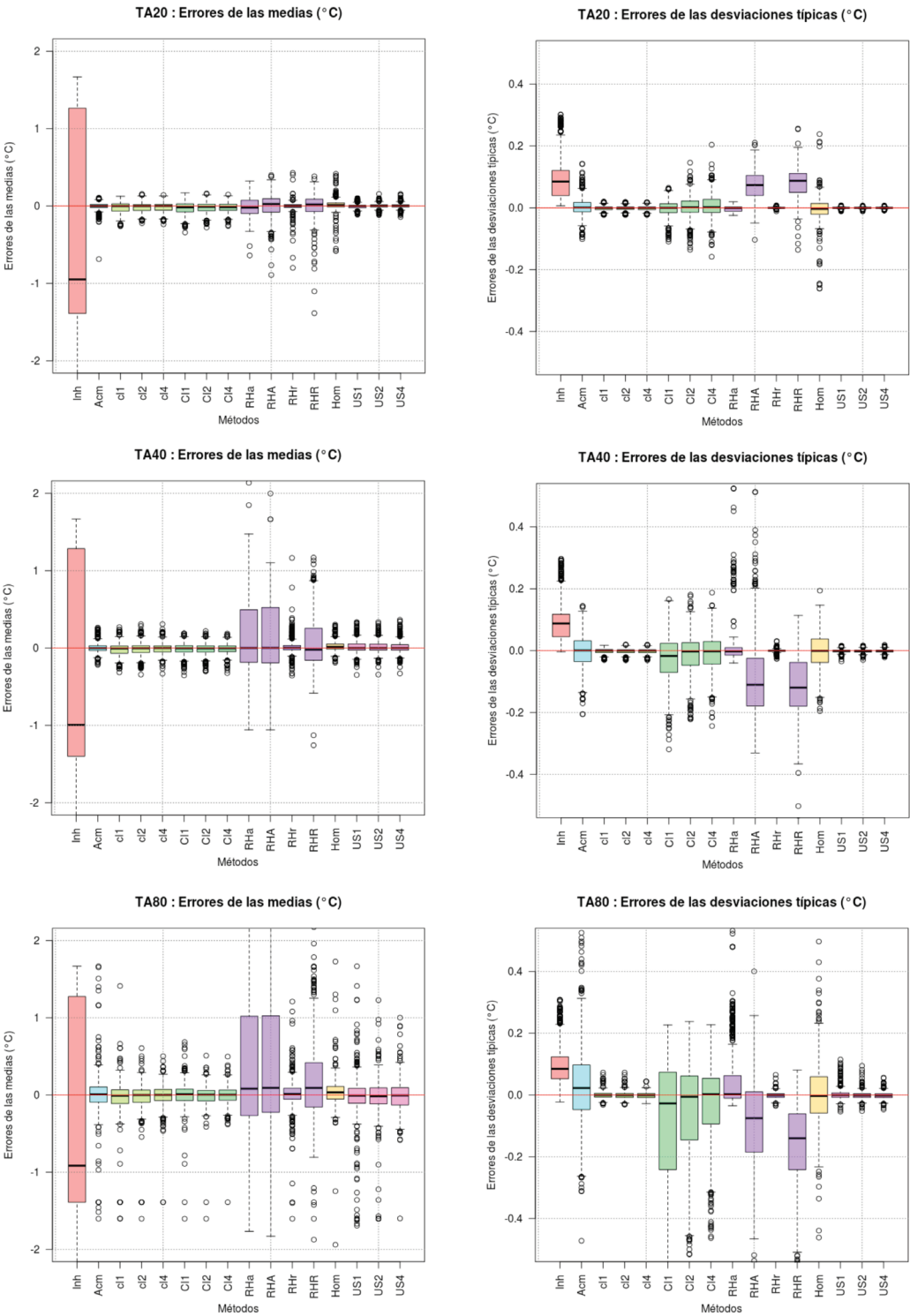


FIG. 5: Errores de las medias (izquierda) y de las desviaciones típicas (derecha) de las series problema, antes (Inh) y después de ser homogeneizadas por los programas ensayados. (De arriba a abajo, redes con correlaciones buenas, regulares y malas).

Respecto al uso de la información adicional proporcionada por las series incompletas, solamente se aprecia una clara mejoría en las redes TA80, que presentan las peores correlaciones interestaciones. Por consiguiente, la elevada tolerancia a ausencia de datos de Climatology y USHCN_v52d es más una ventaja para esas mismas series (que ven sus datos rellenados) que para las series con registros más completos. Las pruebas C11, C12 y C14 se diferencian de las c11, c12 y c14 en que estas últimas aplican una corrección constante, mientras que las primeras ajustan también la desviación típica. Como sabemos de antemano que las series problema presentan saltos de magnitud constante, resulta satisfactorio comprobar que el intento de corregir cambios de variabilidad no empeora mucho los resultados, cosa que sí sucede con las soluciones RHR.

En cuanto a las diferencias entre las tendencias de las series homogeneizadas y sus valores originales homogéneos, los resultados son similares: una vez rechazadas las homogeneizaciones absolutas y las de RHR, los demás muestran valores parecidos, con sesgos despreciables y pequeñas diferencias en cuanto a la dispersión de las soluciones. Las diferencias de medias y desviaciones típicas se presentan en la figura 5. Mientras que las diferencias de medias muestran resultados similares a los de las tendencias, las de las desviaciones típicas tienen algunas peculiaridades: Los métodos que aplican correcciones constantes (c11/2/3, RHR y US1/2/3) proporcionan los mejores resultados, mientras que los ajustes por cuantiles de RHA y RHR dan los más sesgados. C11 y C12 producen valores muy dispersos cuando las correlaciones son bajas (TA80), que mejoran con el uso de la información de series incompletas en C14, aproximando sus resultados a los intermedios proporcionados por Hom y AcM.

Por tanto, una vez eliminadas las homogeneizaciones absolutas, la elección de uno de estos paquetes dependerá un poco del tipo de problema a resolver y de las preferencias del usuario: Si se desea una aplicación interactiva se puede optar por HOMER o RHTestV3, si bien el primero necesita de cierto entrenamiento, y el segundo requiere que se le suministre una serie de referencia adecuada para cada serie problema. Si vamos a trabajar con temperaturas podemos usar ACMANT, que se desarrolló específicamente para esta variable con su claro patrón estacional unimodal, o USHCN_v52d. Éste último se obtiene en código fuente escrito en FORTRAN, y hay que compilarlo en una máquina adecuada, mientras que ACMANT se suministra como programa ejecutable para DOS/Windows (como MASH y AnClim, no probados todavía por la dificultad de su automatización). Climatology no está diseñado únicamente para temperaturas y es de código abierto, pero no precisa de compilación pues, al igual que RHTestV3 y HOMER, está escrito en R, bastando tener instalado este paquete de computación estadística, libre y disponible gratuitamente casi para cualquier sistema operativo. (Una comparativa más exhaustiva de las características y modos de obtener estos y otros paquetes de homogeneización puede consultarse en <http://www.climatology.eu/DARE>).

4. CONCLUSIONES Y TRABAJOS FUTUROS

La comparación entre las soluciones aportadas por cada método y las series originales muestran que, aparte de los esperables malos resultados de la homogeneización absoluta, las demás no presentan grandes diferencias. También se constata que, tal como cabía prever, los errores aumentan cuando la correlación interestaciones disminuye.

La información aportada por las series cortas sólo aporta mejorías claras en la reconstrucción de las series problema cuando las correlaciones son bajas, de modo que el mayor beneficio de incluir esas series de corta duración en el análisis reside en el relleno de sus propias lagunas de datos.

Por otra parte, resulta tranquilizador comprobar que, a pesar de las fuertes inhomogeneidades introducidas en este experimento, las series homogeneizadas no presentan ningún sesgo apreciable en sus tendencias, lo que aumenta la confianza en su uso para la evaluación de la variabilidad del clima.

Se ha comprobado la utilidad de los bancos de datos de prueba automáticos, que permiten la comparación de un gran número de resultados con poco esfuerzo, siempre que los programas de homogeneización permitan su uso en modo no interactivo, por lo que en el futuro se espera poder ampliar la intercomparación a otros métodos automatizables, extender la problemática a valores con distribución sesgada (como la precipitación), e introducir un mayor número de saltos de variada magnitud y con ciclo anual, aumentando así la dificultad de su detección al tiempo que se aproximan a la problemática de las series climatológicas reales.

Agradecimientos

Es de agradecer la ayuda prestada por los desarrolladores cuando han surgido problemas para automatizar sus programas.

BIBLIOGRAFÍA

- Aguilar, E., Auer, I., Brunet, M., Peterson, T.C., Wieringa, J. (2003). *Guidelines on climate metadata and homogenization*. World Meteorological Organization, WMO-TD No. 1186, WCDMP No. 53, Geneva, Switzerland, 55 pp.
- Domonkos, P. (2011). Adapted Caussinus-Mestre Algorithm for homogenising Networks of Temperature series (ACMANT). *Int. J. Geosci.*, 2, pp. 293–309, doi:10.4236/ijg.2011.23032.
- Guijarro, J.A. (2011). *User's guide to climatol*. 40 pp. <http://www.meteobal.com/climatol/climatol-guide.pdf>
- Menne, M.J., Williams, C.N.Jr. y Vose, R.S. (2009). The U.S. historical climatology network monthly temperature data, version 2, B. *Am. Meteorol. Soc.*, 90, pp. 993–1007, doi: 10.1175/2008BAMS2613.1.
- Peterson, T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Førland, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., Parker, D. (1998). Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.*, 18, pp. 1493–1517.
- R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org/>.
- Venema, V., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C.N., Menne, M., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom, M., Likso, T., Esteban, P. y Brandsma, T. (2012). Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8, pp. 89–115.
- Wang, X.L. y Feng, Y. (2010). *RHtestsV3 User Manual*. http://cccma.seos.uvic.ca/ETCCDMI/RHtest/RHtestsV3_UserManual.doc